Preprint version. Please cite original version:

Díaz-Bravo, Rocío y Vaamonde, Gael (2020): «Creación de ediciones digitales para lingüistas de corpus: el caso del Retrato de la Loçana andaluza». En J.R. Belda-Medina y R. Casañ-Pitarch (Eds.): *Análisis del Discurso en la Era Digital: Una Recopilación de Casos de Estudio*, Granada, Editorial Comares, pp. 17-34.

CREACIÓN DE EDICIONES DIGITALES PARA LINGÜISTAS DE CORPUS: EL CASO DEL *RETRATO DE LA LOÇANA ANDALUZA*

Rocío Díaz-Bravo / Gael Vaamonde

rociodiazbravo@ugr.es / gaelvaamonde@ugr.es

Universidad de Granada

1. Introducción

El *Retrato de la Loçana andaluza* (en adelante, RLA) es una novela dialogada escrita en Roma en 1524 y publicada en Venecia hacia 1530. Su autor, el clérigo andaluz Francisco Delicado, pretende realizar un retrato de su protagonista, la prostituta Lozana, así como de los numerosos (139) y variados personajes de la Roma multicultural y plurilingüe anterior al saqueo de 1527 (Díaz-Bravo, 2019: 1). Se trata de una obra literaria del Siglo de Oro excepcionalmente rica en lengua hablada (Anipa, 2001: 8), representativa de la inmediatez comunicativa (Koch y Oesterreicher, 2007) –más concretamente, de lo hablado escrito– y citada por Oesterreicher (2004: 755) como un ejemplo célebre entre los textos literarios áureos caracterizados por la mímesis de lo hablado, pues permite rastrear huellas de oralidad pasada. Aparecen personajes de diferente origen geográfico y social, situaciones comunicativas con distinto grado de formalidad, así como diversas variedades discursivas (Díaz-Bravo, 2010: 225-240). Nos encontramos, por tanto, ante un texto particularmente interesante para la realización de estudios lingüísticos.

El objetivo de este capítulo es presentar *LD. Lozana Digital* (en adelante, LD), un recurso electrónico que combina una edición digital de esta obra y un corpus anotado, especialmente diseñado para analizar el texto desde un punto de vista lingüístico. Este recurso (Díaz-Bravo y Vaamonde, 2019) favorece la realización de análisis lingüísticos gracias a la combinación de tres aspectos fundamentales: (i) una edición crítica digital del texto, (ii) una base de datos con las semblanzas de los personajes, y (iii) un corpus normalizado, lematizado y anotado. Estos tres aspectos han sido implementados mediante el uso de estándares ya consolidados en el campo de las Humanidades Digitales y la Lingüística de Corpus, respectivamente: el estándar del consorcio TEI (*Text Encoding Initiative*) se ha usado para la edición digital y la base de datos, y el estándar EAGLES (*Expert Advisory Group on Language Engineering Standards*) de lenguas europeas, para las etiquetas morfosintácticas del corpus.

La creación de LD se ha realizado a través de TEITOK, una plataforma en línea pensada para construir corpus electrónicos que combinan marcación textual con anotación lingüística.

2. Estado de la Cuestión

El único ejemplar impreso[1] antiguo del RLA que ha sobrevivido hasta nuestros días se encuentra en Viena, en la Biblioteca Nacional de Austria (en adelante, BNA), donde fue encontrado en 1845 por el romanista Ferdinand Wolf, después de más de tres siglos de olvido y de absoluto silencio. La publicación de una edición facsimilar a mediados del siglo pasado (Pérez-Gómez, 1950), basada en el ejemplar de la BNA, fue el punto de partida para que esta obra se diera a conocer, gracias a la sucesión de ediciones (Damiani, 1969; Allegra, 1983; Allaigre, 1985; Chiclana, 1988; Gernert y Joset, 2013, entre otros). Dicha edición facsimilar está disponible en línea en la Biblioteca Virtual Miguel de Cervantes (en adelante, BVMC) desde 2003, como una colección de imágenes que se pueden visualizar en pantalla y descargar en formato JPEG. Asimismo, desde 2016, es posible descargar en formato JPEG o consultar en línea las imágenes del único ejemplar supérstite –a través de la página electrónica de la BNA–, con opciones de ampliación de la imagen que permiten visualizarla con alta calidad y resolución. Esto es especialmente útil para fragmentos que no están claros en el facsímil:

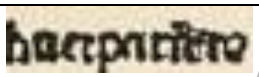
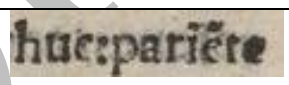
Edición facsimilar disponible en línea en BVMC	Edición facsimilar disponible en línea en BNA
	

Tabla 1. Fragmento *hue pariente* en dos ediciones facsimilares (folio 23r)

Por otra parte, la BVMC también ofrece dos ediciones del texto en formato HTML. Una de ellas (BVMC, 2004) se presenta como una "edición modernizada de *La lozana andaluza*" basada en la edición de Damiani (1984 [1969]), en la que "las marcas de editor" se han eliminado "con el objetivo de facilitar la lectura del texto al público no especializado", como se advierte en nota preliminar. La otra versión (BVMC, 2003) se basa en "los criterios ortográficos y de puntuación utilizados por Chiclana en su edición crítica" (1988), aunque se indica que los correctores de la BVMC "han introducido un total de 43 variantes, casi todas ellas erratas subsanadas".

En ambos casos nos encontramos ante ediciones y no corpus anotados, lo que implica que no son recursos útiles para el análisis lingüístico del texto. No obstante, ambas permiten obtener resultados de búsquedas en forma de concordancias (no lematizadas), como se observa a continuación (figuras 1 y 2).

[1] No contamos con ninguna copia manuscrita del siglo XVI.

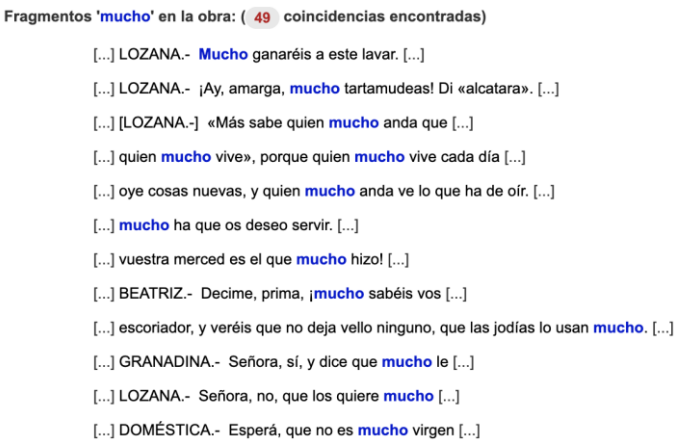


Figura 1. Concordancias de *mucho* en BVMG (2004)

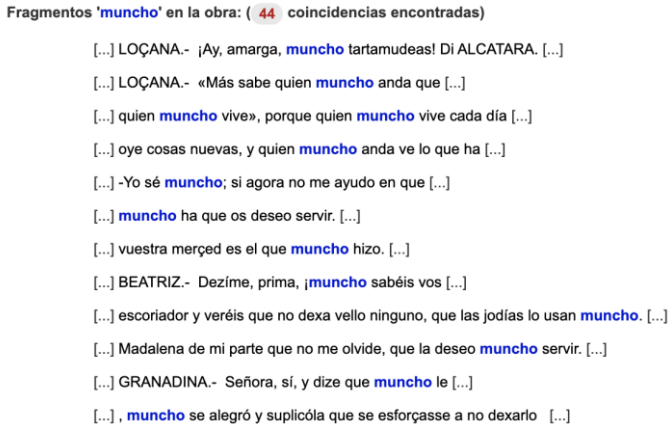


Figura 2. Concordancias de *muncho* en BVMG (2003)

La edición basada en Chiclana (1988) constituye la versión con grafía más conservadora de las dos que encontramos en el BVMC[2]. Es esta la que ha sido incorporada al CdE[3]. Por otra parte, tanto el CORDE como el CDH de la RAE recogen el texto de la edición de Allaigre (1985), en concreto, una reedición de 1994. Asimismo, a partir de la misma edición, se ha elaborado un "Glossario della *Lozana andaluza*" con la traducción de todos los lemas al italiano, organizado en orden alfabético en archivos PDF (Tempesta y Carocci). Díaz-Bravo (2010), en su Tesis Doctoral publicada en línea en formato PDF, preparó las concordancias lematizadas del RLA, creadas a partir de su edición crítica.

Finalmente, algunas de las últimas ediciones del RLA se pueden leer como libro electrónico (Bubnova, 2008; Díaz-Bravo, 2019). Estos formatos (PDF, ePub…) son útiles para la lectura del texto, pero no para realizar un procesamiento automático que facilite su análisis lingüístico.

[2] Por eso encontramos, como se observa en la figura 2, el mantenimiento de la epéntesis nasal en *muncho*: un rasgo característico de Delicado como hablante andaluz y, muy posiblemente, judío.

[3] Por cierto, en CdE *La Lozana andaluza* aparece catalogada erróneamente como un texto de 1510.

3. Metodología

Para la creación del recurso electrónico que aquí presentamos se han utilizado cuatro fuentes de datos. La tabla 2 resume las características de estas cuatro fuentes, que constituyen el punto de partida en LD.

	Descripción	*Formato*
1	transcripción del texto (Díaz-Bravo, 2019)	TXT
2	transcripción del texto (Díaz-Bravo, 2010)	TXT
3	hoja de datos con información de los personajes	Microsoft Excel
4	conjunto de imágenes facsimilares	JPEG

Tabla 2. Conjunto de datos utilizados para la creación de LD

En primer lugar, se ha hecho uso de la edición crítica del *Retrato de la Loçana andaluza* realizada por Díaz-Bravo (2019); en concreto, se ha tomado como punto de partida un archivo de texto simple en formato TXT que contiene únicamente la transcripción del texto tal como aparece en dicha edición, esto es, sin el aparato crítico ni el estudio lingüístico que acompaña a la edición impresa. Esta transcripción sigue los siguientes criterios de edición: se conservan las grafías originales y el uso de arcaísmos (ej.: *delantre, labrios*), variantes diatópicas (ej.: *meatad*) y variantes diastráticas (ej.: *açiprés*); se moderniza la acentuación, el uso de mayúsculas y minúsculas, la puntuación original, la tilde nasal y las abreviaturas, que se transcriben ya desarrolladas (Díaz-Bravo, 2019: 23-25); los fragmentos de texto añadidos por el editor –generalmente la entrada de personajes en determinados pasajes– aparecen con la marca <> (*vid.* figura 3). Finalmente, no se conservan los cambios de párrafo ni la separación de líneas del texto original, pues se trata de una edición que pretende facilitar la lectura.

En segundo lugar, se ha utilizado un archivo TXT con los mismos criterios de edición citados anteriormente, pero con la particularidad de conservar los cambios de párrafo y la separación de líneas del texto original (Díaz-Bravo, 2010)[4].

En tercer lugar, se ha hecho uso de un archivo creado en Microsoft Excel que fue utilizado en su momento para la investigación recogida en Díaz-Bravo (2010) y que contiene información variada acerca de cada personaje incluido en la obra (edad, oficio, procedencia geográfica…).

En cuarto y último lugar, se ha hecho uso de la edición facsimilar del RLA disponible en la página electrónica de la BNA. Este conjunto de imágenes en formato JPEG ha sido descargado e importado a la plataforma TEITOK para su visualización[5].

El trabajo que aquí presentamos es resultado de un procesamiento de las tres primeras fuentes de datos junto con la importación y vinculación al texto del conjunto de imágenes en JPEG. El objetivo final es la creación de un único recurso electrónico que combina una edición digital del *Retrato de la Loçana andaluza* y un corpus anotado para el análisis lingüístico de la obra. Los pasos que se han seguido para alcanzar este objetivo son, por orden de aplicación, los que se indican a continuación:

[4] La edición crítica presentada en la Tesis Doctoral de Díaz-Bravo (2010) carece, no obstante, del trabajo de revisión adicional realizado posteriormente en Díaz-Bravo (2019).

[5] Agradecemos a la BNA que nos haya concedido permiso para hacer uso de las imágenes en LD.

a) Fusión de las dos transcripciones TXT en una única versión
b) Conversión de dicha versión fusionada a lenguaje TEI-XML
c) Creación de fichas de personajes en lenguaje TEI-XML
d) Importación y vinculación de imágenes
e) Tokenización del texto
f) Normalización ortográfica del texto
g) Anotación morfosintáctica y lematización

La aplicación de estos pasos se ha realizado adoptando una estrategia semiautomática, esto es, un procesado automático de los datos con revisión manual del resultado en determinadas fases del proceso. Los tres primeros pasos (a-c) constituyen una tarea de preprocesamiento de los datos de partida y se han realizado usando *scripts* basados en lenguaje Perl. Los cuatro pasos restantes (d-g) fueron implementados a través de la plataforma web TEITOK (Janssen, 2014), que explicamos más abajo; los tres últimos pasos (e-g) están centrados en el procesamiento lingüístico del texto, esto es, la creación del corpus anotado.

3.1. Fusión de las dos Transcripciones

Uno de los objetivos que nos planteamos en la creación de LD fue ofrecer una edición digital lo más conservadora posible a partir de nuestros datos de partida. Así, LD ofrece el texto plano de la edición de Díaz-Bravo (2019), pero con los cambios de línea y párrafo de la obra original, como en Díaz-Bravo (2010). Para ello, se ejecutó un *script* que leyó ambas versiones y creó de manera automática una versión fusionada, sobre la que se aplicó una revisión manual. Se ofrece a continuación (figura 3) un fragmento de dicha versión:

```
Mamotreto II. Responde la tía y prosigue

<Tía>: Sobrina, más ha de los años treynta que io no vi a vuestro padre, porque se fue
niño; y después me dixeron que se casó por amores con vuestra madre, y en vos
veo io que vuestra madre hera hermossa. Loçana: ¿Yo, señora? Pues más paresco
a mi agüela que a mi señora madre; y por amor de mi agüela me llamaron a mí
Aldronça; y si esta mi agüela biuía, sabía yo más que no sé, que ella me mostró […]
```

Figura 3. Fragmento de texto simple usado como fuente de datos para LD

3.2. Conversión del Texto a Lenguaje TEI-XML

Una vez fusionadas las dos transcripciones de partida en un único archivo TXT, el siguiente paso consistió en convertir los datos a un lenguaje que asegure su preservación a largo plazo, facilite su visualización en línea y amplíe las opciones de búsqueda y de recuperación de información. Actualmente, ese lenguaje es XML (*eXtensible Mark-up Language*):

> […] the goals of corpus linguistics and language documentation are not so different. Both fields aim for collections of related language data that are interoperable, searchable, reusable, and mobilizable for a broad range of linguistic inquiry […]. Current advances in encoding and interoperability like XML and Unicode are already making this possible (Gries y Berez, 2017: 404).

En consonancia con las prácticas actuales en el campo de las Humanidades Digitales, la conversión del corpus a lenguaje XML se ha realizado adoptando los estándares de codificación propuestos por el consorcio TEI para la edición de textos en formato digital (TEI Consortium, 2019a). Para este propósito, se ha creado un *script* diseñado específicamente para añadir marcas TEI-XML de forma automática. A modo de ejemplo, ofrecemos en la figura 4 el mismo fragmento de la figura 3, ahora con marcación TEI-XML:

```
<milestone type="section" n="Mamotreto II"/>
   <p>
      <lb/> Mamotreto II. Responde la tía y prosigue
   </p>
   <p>
      <lb/> <speaker><supplied reason="ommited">Tía:</supplied></speaker>
   </p>
   <p>
      <lb/> <sp n="Tía (de Lozana)" who="#tia1" ><hi rend="capital">S</hi>obrina, más ha de los
      años treynta que io no vi a vuestro padre, porque se fue
      <lb/> niño; y después me dixeron que se casó por amores con vuestra madre, y en vos
      <lb/> veo io que vuestra madre hera hermossa.</sp> <sp n="Lozana">
      <speaker>Loçana:</speaker> ¿Yo, señora? Pues más paresco
      <lb/> a mi agüela que a mi señora madre; y por amor de mi agüela me llamaron a mí
      <lb/> Aldronça; y si esta mi agüela biuía, sabía yo más que no sé, que ella me mostró
      [...] </sp>
   </p>
```

Figura 4. Fragmento de texto en TEI-XML

Como se deduce de la comparación de las figuras 3 y 4, el *script* de conversión a TEI-XML ejecuta una serie de sustituciones sobre el texto simple basadas en la adición de elementos TEI, como son <p> para la marca de párrafo, <lb/> para el cambio de línea, <milestone/> para el inicio de cada sección textual, <supplied> para las intervenciones del editor, <sp> para la intervención de cada personaje y <speaker> para el nombre que da entrada a cada intervención. Además, se ha añadido manualmente el elemento <hi> para marcar la presencia de letras capitales en el texto original.

La adopción del estándar internacional TEI asegura la integración de LD con otros recursos y repositorios digitales; además, permite mejorar la visualización de la edición digital (*vid.* figura 6) y aumentar las posibilidades de búsqueda (*vid.* apartado 4).

3.3. Creación de Fichas de Personajes

El tercer punto de partida que se ha utilizado para la creación de RLA es un archivo con información variada sobre cada personaje de la obra. Este archivo también fue convertido automáticamente –desde su formato inicial en Microsoft Excel– a lenguaje TEI-XML para facilitar su procesamiento. El resultado de esta conversión es un archivo XML con la información de cada personaje debidamente marcada y catalogada para su posterior vinculación con los datos textuales.

La ficha de cada personaje aparece recogida mediante un elemento <person>, que contiene a su vez diferentes elementos XML en función del tipo de información recogida (TEI Consortium, 2019b). Esta información incluye cuestiones relativas al nombre del personaje en la obra (nombre original, nombre normalizado, nombres alternativos), la lista de mamotretos (capítulos) en que interviene o aspectos de carácter puramente social (edad,

religión, ocupación, educación, procedencia…). Recogemos en la figura 5 un ejemplo de ficha biográfica tomado del personaje *Tía (de Lozana)*.

```
<person sex="mujer" role="hablante" xml:id="tia1">
    <persName>
        <name type="original">Tía</name>
        <name type="normalized">Tía (de Lozana)</name>
    </persName>
    <affiliation>tía de Loçana</affiliation>
    <age>adulto</age>
    <education>analfabeto</education>
    <faith/>
    <index>II, III</index>
    <nationality>Sevilla</nationality>
    <residence type="italy">no residente</residence>
    <occupation>alcahueta</occupation>
    <socecStatus type="stratum">no privilegiado</socecStatus>
    <socecStatus type="economical">pobre</socecStatus>
    <socecStatus type="social">no prestigioso</socecStatus>
    <trait>
        <desc>blanca</desc>
    </trait>
</person>
```

Figura 5. Ficha de personaje en TEI-XML

Cada personaje está asociado a un identificador único a través del valor del atributo @xml:id (*tia1* en el ejemplo anterior). Este identificador se repite como valor del atributo @who en cada una de las intervenciones del personaje en el propio texto (*vid.* figura 4). Se establece de esta forma un vínculo entre la información de cada personaje y su discurso en el texto, abriendo así la posibilidad de realizar búsquedas cruzadas que faciliten un análisis de la obra desde una perspectiva sociolingüística (*vid*. figura 13).

3.4. Importación y Vinculación de las Imágenes

La cuarta fuente de datos utilizada en LD es la edición facsimilar de la obra, descargable como conjunto de imágenes en formato JPEG desde la página electrónica de la BNA. Este conjunto de imágenes –una por página– fue importado a la plataforma TEITOK para ofrecer así una visualización alternativa al texto que facilite la consulta de cualquier aspecto de la obra en el documento original.

Además de importar las imágenes, se creó un vínculo entre texto e imagen con el objetivo de ofrecer de manera simultánea tanto la visualización del texto como la visualización de la imagen facsimilar correspondiente. Dicho vínculo se creó a través del elemento <pb/>, que es el elemento propuesto en TEI para marcar inicio de página. Este elemento incluye un atributo @facs, cuyo valor se recuperó secuencialmente a partir del nombre del archivo de imagen JPEG correspondiente (1r.jpg, 1v.jpg, 2r.jpg…). De esta forma, el sistema reconoce la correspondencia entre texto e imagen y ambos pueden ser visualizados en pantalla de forma paralela, como se aprecia en la figura 6:

Sobrina, más ha de los años treynta que io no vi a vuestro padre, porque se fue niño; y después me dixeron que se casó por amores con vuestra madre, y en vos veo io que vuestra madre hera hermossa. Loçana: ¿Yo, señora? Pues más paresco a mi agüela que a mi señora madre; y por amor de mi agüela me llamaron a mí Aldronça; y si esta mi agüela biuía, sabía yo más que no sé, que ella me mostró

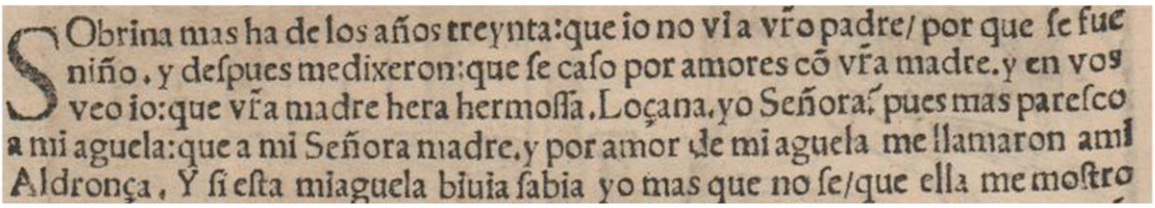


Figura 6. Fragmento visualizado como texto e imagen[6]

Finalmente, mediante uso de lenguaje CSS se personalizaron los estilos de visualización del texto, bien para representar lo más fielmente posible la impresión del original (ej.: creación de letras capitales), bien para facilitar la lectura (ej.: marcar en azul la entrada de cada personaje).

3.5. Procesamiento Lingüístico

Los tres primeros pasos explicados en este apartado están dirigidos a convertir los datos de partida en TEI-XML, encauzando así uno de los objetivos de LD, que es la creación de una edición digital. El otro objetivo es convertir la edición en un corpus anotado que permita realizar estudios lingüísticos. Ambos productos, el de la edición digital y el del corpus anotado, deben ser idealmente combinados en un recurso que facilite al usuario tanto la visualización del texto como la búsqueda en el corpus. En el caso de LD, ese recurso es TEITOK.

> TEITOK is a web-based system for viewing, creating, and editing corpora with both rich textual mark-up and linguistic annotation. For visitors, the system provides a graphical user interface in which the annotated document can be visualized in a number of different ways, depending on what the user is interested in. And for administrators of the corpus, TEITOK uses the same interface to easily and efficiently edit the underlying XML document (Janssen, 2014).

La funcionalidad de TEITOK es doble: por un lado, permite al usuario navegar a través de la edición digital y, simultáneamente, realizar búsquedas en el corpus anotado; por otro lado, permite al administrador procesar y editar los datos en XML. Respecto a esta segunda funcionalidad, TEITOK incorpora un tokenizador, un normalizador ortográfico y un anotador morfosintáctico que permiten procesar el texto y enriquecerlo con información lingüística para facilitar y ampliar las posibilidades de análisis posterior (*vid*. apartado 4). Estos tres pasos se realizan de forma consecutiva y automática, si bien es necesaria una tarea de revisión manual sobre el resultado del normalizador ortográfico y del anotador morfosintáctico.

[6] Por razones de espacio, se ofrece aquí la imagen debajo del texto; en LD, la imagen se visualiza a la derecha del texto.

La utilidad de TEITOK para la crear corpus lingüísticos a partir de ediciones digitales ha sido demostrada en proyectos como *ODE* (Calderón-Campos, en prensa) o *Post Scriptum* (Vaamonde 2018a, 2018b), entre otros.

3.5.1. Tokenización

El proceso de conversión de la edición digital en un corpus lingüístico se inicia con la tokenización, esto es, la marcación de cada palabra y de cada signo de puntuación del texto. Este proceso se realiza automáticamente. TEITOK crea un elemento <tok> por cada *token* y le asigna un identificador único, como se recoge en la figura 7[7]:

```
<tok id="w-1440">y</tok>
<tok id="w-1441">en</tok>
<tok id="w-1442">vos</tok>
<lb/>
<tok id="w-1443">veo</tok>
<tok id="w-1444">io</tok>
<tok id="w-1445">que</tok>
<tok id="w-1446">vuestra</tok>
<tok id="w-1447">madre</tok>
<tok id="w-1448">hera</tok>
<tok id="w-1449">hermossa</tok>
<tok id="w-1450">.</tok>
```

Figura 7. Fragmento con marcación de *tokens*

3.5.2. Normalización Ortográfica

Sobre el texto tokenizado, se pasa un normalizador ortográfico que, también de manera automática, asocia a cada forma transcrita su correspondiente forma con grafía normalizada de acuerdo con la ortografía estándar actual.

```
<tok id="w-1440">y</tok>
<tok id="w-1441">en</tok>
<tok id="w-1442">vos</tok>
<lb/>
<tok id="w-1443">veo</tok>
<tok id="w-1444" nform="yo">io</tok>
<tok id="w-1445">que</tok>
<tok id="w-1446">vuestra</tok>
<tok id="w-1447">madre</tok>
<tok id="w-1448" nform="era">hera</tok>
<tok id="w-1449" nform="hermosa">hermossa</tok>
<tok id="w-1450">.</tok>
```

Figura 8. Fragmento con normalización ortográfica

[7] El elemento <tok> no está contemplado por el estándar TEI, que sí contempla un elemento tokenizador <w>. En Janssen (2016: 4038) se explican las ventajas de utilizar <tok> en lugar de <w>.

La información relativa a la normalización ortográfica se marca como valor de un atributo @nform dentro de cada *token*. Siguiendo con el ejemplo anterior, las formas *io*, *hera* y *hermossa* se asociarán a las formas *yo*, *era* y *hermosa*, respectivamente (*vid*. figura 8). La finalidad de este proceso es triple. En primer lugar, se obtiene una versión que facilita la lectura del texto al público lego. En segundo lugar, se amplían las opciones de búsqueda en el corpus, ya que es posible buscar por forma original o por forma normalizada; además, la utilización de ambos niveles permite agrupar los resultados de una misma forma normalizada en función de sus formas transcritas asociadas, lo que facilita el estudio de la variación gráfica en el texto (*vid*. tabla 3). En tercer lugar, se facilita el acierto del anotador morfológico, que no necesita etiquetar las formas hipotéticas *hermosa*, *hermossa*, *ermosa*, *ermossa*, sino únicamente la forma *hermosa*.

3.5.3. Anotación Morfosintáctica

Por último, un anotador automático asocia cada palabra contenida en el texto a un lema (@lemma) y una etiqueta morfosintáctica (@pos), y el resultado de esa anotación es revisado manualmente. El sistema de etiquetas utilizado para anotar el corpus de LD está basado en la propuesta del grupo EAGLES para la anotación morfosintáctica de lexicones y corpus para todas las lenguas europeas (Leech y Wilson, 1996). El conjunto de etiquetas EAGLES se rige por un sistema de posiciones: cada etiqueta consta de una secuencia de letras y números, donde cada letra o número representa un rasgo morfosintáctico determinado dependiendo de su posición dentro de la secuencia.

Así, por ejemplo, la forma *hermosa* está asociada al lema *hermoso* y a la etiqueta AQFS00, que representa adjetivo (A), calificativo (Q), femenino (F), singular (S). De modo análogo, la forma *veo* está asociada al lema *ver* y a la etiqueta VMIP1S0, esto es, verbo (V), principal (M), indicativo (I), presente (P), primera persona (1), singular (S)[8]. Se recoge en la figura 9 el fragmento anterior una vez anotado y lematizado.

```
<tok id="w-1440" lemma="y" pos="CC">y</tok>
<tok id="w-1441" lemma="en" pos="S">en</tok>
<tok id="w-1442" lemma="vos" pos="PP2CS0">vos</tok>
<lb/>
<tok id="w-1443" lemma="ver" pos="VMIP1S0">veo</tok>
<tok id="w-1444" nform="yo" lemma="yo" pos="PP1CSN">io</tok>
<tok id="w-1445" lemma="que" pos="CS">que</tok>
<tok id="w-1446" lemma="vuestro" pos="DP2FSP">vuestra</tok>
<tok id="w-1447" lemma="madre" pos="NCFS000">madre</tok>
<tok id="w-1448" nform="era" lemma="ser" pos="VSII3S0">hera</tok>
<tok id="w-1449" nform="hermosa" lemma="hermoso" pos="AQFS00">hermossa</tok>
<tok id="w-1450" lemma="." pos="Fp">.</tok>
```

Figura 9. Fragmento con lematización y anotación morfológica

Como se aprecia en la figura 9, el resultado final de este proceso es un archivo XML en donde, para cada *token*, se recoge la forma transcrita (*i.e.* texto marcado con el elemento <tok>) y las correspondencias oportunas relativas a los diferentes niveles de edición (*i.e.* valores de atributos dentro del elemento <tok>). Este documento XML es convertido

[8] El etiquetario completo utilizado en LD está disponible en la siguiente URL: <http://corpora.ugr.es/lozana/index.php?action=tagset>

automáticamente en un corpus sobre el que se pueden realizar búsquedas de diferente tipo directamente desde la plataforma en línea, tal como se explica en el apartado siguiente.

4. Resultados y Análisis

TEITOK incorpora un sistema de búsqueda desde el que se pueden realizar diferentes consultas en el corpus. Las consultas se pueden expresar directamente en lenguaje CQP (*Corpus Query Processor*) o se pueden construir a través de una interfaz intuitiva que convierte las consultas del usuario en lenguaje CQP. En LD, esta interfaz está dividida en dos bloques: búsqueda en el texto y búsqueda por personaje (*vid.* figura 10). Dedicamos este apartado a ilustrar algunas de las múltiples posibilidades de explotación del corpus LD a través de dicha interfaz.

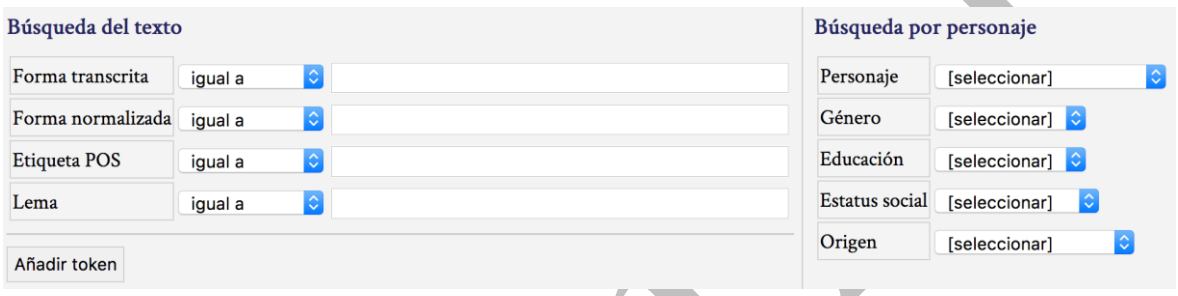


Figura 10. Interfaz de consultas

Si nos interesa buscar la forma *cassa* con *s* doble –que originalmente representaba un sonido sordo, frente a *s* simple, que representaba un sonido sonoro, como en italiano–, tendríamos que realizar una búsqueda por "forma transcrita". Como es habitual en Lingüística de Corpus, el sistema devuelve los resultados en formato KWIC (*Key Word in Context*), es decir, Palabra Clave en Contexto (PCEC), así como en formato de contexto ampliado:

contexto	tarde, mirá que es vna **cassa** nueua pintada y dos gelossías |	y
contexto	busque vna persona que mire por **cassa** , pues que ni vuestra merçed	
contexto	vez. Y ella tiene su **cassa** por sí, y quanto le	
contexto	haze enbaxadas y mete de su **cassa** muncho almazén, y sábele dar	
contexto	para esso tienpo ay, y **cassa** tengo, que no lo tengo	
contexto	criado, que es ydo a **cassa** , y díxele que truxese dos	
contexto	!, | que pasé por su **cassa** y sospeché que no estaua allí	
contexto	sí, que yo voy a **cassa** de la señora Velasca | para que	
contexto	, que quiero yr a mi **cassa** y, si es venido mi	
contexto	él porfiar y con todas se **cassa** y a ninguna sirue de buena	
contexto	, hi, hi! Vuestra **cassa** buscamos y si no os encontráuamos	
contexto	. Bien que yo y mi **cassa** seamos pobres, al menos	

Figura 11. Concordancias de la forma transcrita *cassa*. CQP: [form="cassa"]

Asimismo, es posible realizar una búsqueda de la forma normalizada de la palabra *casa*, que incluirá tanto las variantes transcritas con *s* como las que han sido transcritas con *ss*:

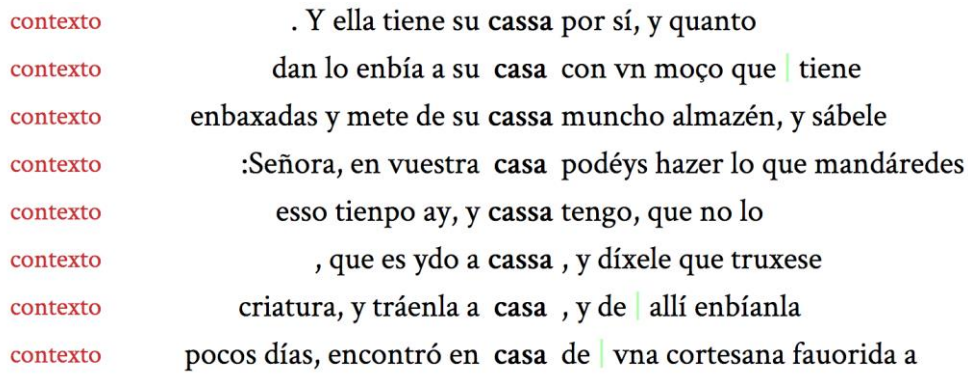


Figura 12. Concordancias de la forma normalizada *cassa*. CQP: [nform="casa"]

Por otra parte, los resultados de la búsqueda de una forma normalizada se pueden visualizar de manera agrupada y ordenada –de mayor a menor frecuencia o viceversa– en una tabla en la que es posible comparar las diferentes formas transcritas asociadas a una misma forma normalizada, lo cual permite realizar estudios de variación ortográfica (por ejemplo: *s-ss*, *h-∅*, *c-ç-z*, *u-v*):

Grupo de formas (*casa*)	Número de ocurrencias
casa	227
cassa	12

Grupo de formas (*hacer*)	Número de ocurrencias
hazer	143
hacer	4
haçer	1
azer	1

Grupo de formas (*había*)	Número de ocurrencias
auía	38
hauía	8
avía	1

Tabla 3. Grupos de formas transcritas asociadas a una misma forma normalizada

Como el corpus incluye lemas y etiquetas PoS, las posibilidades de búsqueda se multiplican, como mostramos en los siguientes ejemplos (véanse las figuras 13, 14 y 15, y la tabla 4). Se puede realizar una búsqueda simple por lema. La diversidad de formas es especialmente llamativa en un verbo que además incluye variación ortográfica, como *traer*:

contexto	tengo. tía:Esperá, |	**traeré**	aquel pelador o escoriador y veréys
contexto	, y diré yo que lo	**traygo**	de Leuante. | ranpín:Sea
contexto	valéys que pensáys! | Vamos a	**traer**	vn ganapán que lleue todo esto
contexto	está allí, y dize que	**traxo**	a su hija virgen a Roma
contexto	? ¿Y su madre la	**traxo**	a Roma? ranpín:Señor
contexto	Yo tomé mis dineros, y	**traygo**	vn marauedí de plomo, y
contexto	: "¿Qué mandáys que	**trayga**	?" Loçana: "Vna
contexto	viene otro día cargada, e	**traxo**	| otros dos julios, y metió
contexto	:Pues, ¿quién la	**traxo**	? ranpín:Viene a pleytear
contexto	Válgala, y qué trato que	**trae**	con las manos! Paresçe que
contexto	. ¿Qué es aquello que	**trae**	? Demandémoselo. ¿Qué priesa
contexto	a cassa, y díxele que	**truxese**	dos coxines vazíos para lleuar faxadores

Figura 13. Búsqueda simple por lema. CQP: [lemma="traer"]

Dentro de un plano meramente lingüístico, podemos complicar la búsqueda si nos interesa realizar, por ejemplo, un análisis de carácter fonético-fonológico sobre la variación de las vocales velares átonas *o-u*[9] en el verbo *cubrir* y sus derivados. Para ello, tendríamos que indicar que el lema "termina en" *cubrir*; de esta manera, se incluirá el infinitivo combinado con diversos prefijos:

contexto	mí todas essas cosas?	**Descubrí**	, que lo sirua yo
contexto	chica fossa en diez días |	**cobriste**	y encerraste, dando | fin
contexto	, que, aunque se	**cubra**	, que no aprouecha,
contexto	y por ver si se	**cubriera**	. Mas no curéys,
contexto	vn su vestido | que se	**cubriese**	. Y viéndose sola y
contexto	, y no se osan	**descobrir**	, que no vean el
contexto	miedo que yo jamás lo	**descubra**	. porfirio:Señora,
contexto	de mi preterido criado me	**descubrirá**	, porque | ella misma le
contexto	y mal | para quien lo	**descubrió**	. Ermano, ya es
contexto	los dedos, por las	**encobrir**	. seuillana: | ¡Mostrad
contexto	quanto ternán a quien las	**encubra**	y a quien las quiera
contexto	hazéys, y esta libertad	**encubre**	munchos males. ¿Pensáys
contexto	, los moços mismos os	**encubren**	, y tal casa de

Figura 14. Búsqueda con un interés fonético. CQP: [lemma=".*cubrir"][10]

[9] Dicha alternancia vocálica es un "rasgo peculiar de la lengua estándar de la época, como siempre en un grupo reducido de palabras" (Medina-Morales, 2005: 91).

[10] La expresión regular .***cubrir* recupera *cubrir* precedido de cero o más caracteres, es decir, en términos morfológicos, sin o con prefijo.

Otro ejemplo de búsqueda limitada al plano lingüístico consiste en recuperar todos los sustantivos comunes (en la interfaz de búsqueda: PoS empieza por NC = nombre común) y ordenar los resultados por frecuencia de lemas en la pestaña desplegable de frecuencia (*vid.* tabla 4), lo cual nos permite analizar el discurso de la obra mediante la identificación de los temas predominantes. Si excluimos la palabra comodín *cosa*, observamos que los nombres más frecuentes en el RLA están conectados con el ámbito doméstico (*casa*) –muchos diálogos, de carácter privado e íntimo, tienen lugar en una casa–, la vida diaria (*vida, día, tiempo*), las relaciones entre personas y familiares (*señora, merced, mujer, hombre, criado, madre, hijo*) y la prostitución –*puta* ocupa la quinta posición, debido no solo al oficio de la protagonista, sino también a la extensa enumeración de tipos de prostitutas por parte del Valijero. Todo ello concuerda con los temas más relevantes en la obra, donde encontramos numerosas "conversaciones a través de las cuales sabemos de la vida cotidiana de Roma (alquiler de una casa, compras, cocina, amores, prostitución, descripciones de la ciudad de Roma y de las personas)" (Díaz-Bravo y Fernández-Alcaide, 2018: 360). Asimismo, encontramos otras palabras relevantes desde el punto de vista literario: *autor* –debe tenerse en cuenta que el autor tiene un papel fundamental en la obra, ya que aparece como un personaje a lo largo de la misma– y *mamotreto* –nombre usado para referirse a cada uno de los 66 capítulos en los que se divide el Retrato.

Posición según frecuencia	Nombre común (lema)	Número de ocurrencias
1	casa	198
2	cosa	175
3	señora	173
4	merced	161
5	puta	150
6	vida	109
7	mujer	93
8	día	82
9	tierra	81
10	mano	80
11	hombre	80
12	tiempo	69
13	criado	67
14	autor	62
15	madre	60
16	hijo	59
17	manera	57
18	mamotreto	53

Tabla 4. Sustantivos comunes más frecuentes agrupados por lemas[11]

Por último, el hecho de que los personajes hayan sido etiquetados según diversas categorías, posibilita cruzar datos lingüísticos con datos sociales, permitiendo así análisis de tipo sociolingüístico. Podríamos recuperar, por ejemplo, todos los fragmentos hablados por un personaje, o la lista de sustantivos más frecuentes usados por hombres y por mujeres, o el sufijo superlativo *-ísimo* en personajes cultos (*vid.* figura 15) –uno de los

[11] La lematización se ha realizado de manera automática y aún debe revisarse manualmente, por lo que estos resultados deben tomarse con la debida cautela.

rasgos caracterizadores de los personajes con un nivel educativo superior en el RLA (Díaz-Bravo, 2019: 19-20).

contexto	fortalezas, vna en la	**altíssima**	peña y otra dentro en
contexto	dos ermanos Carauajales, ombres	**animosíssimos**	, acusados falsamente de tiranos
contexto	? siluano:Porque su	**castíssima**	madre y | su cuna fue
contexto	traen | el orígine de las	**castíssimas**	romanas, donde munchas y
contexto	qual | allí miraculosamente mató vn	**ferocíssimo**	serpiente, el qual deuoraua
contexto	grandeza en abundançia. Esta	**fortíssima**	peña es tan alta que
contexto	se llama la solícita y	**fortíssima**	y | santíssima Martha, huéspeda
contexto	, el tenplo lapídeo y	**fortíssima**	ara de Marte fue y
contexto	presente consagrado a la	**fortíssima**	santa Marta, donde los
contexto	; por tanto, el	**fortíssimo**	Marte dedicó a este elemento
contexto	dicha capilla | | los huesos de	**fortíssimos**	reyes y animossos maestres de
contexto	la solícita y fortíssima y |	**santíssima**	Martha, huéspeda de Christo

Figura 15. Búsqueda avanzada que combina datos lingüísticos y sociales

5. Discusión

Para poder apreciar mejor las ventajas de LD resulta clarificador indicar brevemente las vías que existen actualmente para realizar estudios lingüísticos sobre el texto del RLA. Estas vías nos llevan necesariamente a acudir a corpus históricos que incluyan este texto y, hasta donde sabemos, los únicos son los grandes corpus diacrónicos de referencia (CdE, CORDE y CDH). En los tres corpus citados, es posible realizar búsquedas y obtener concordancias sobre un único texto, por tanto, en todos ellos es posible ceñirse a la información textual del RLA. Sin embargo, el usuario encontrará limitaciones[12] desde el momento en el que esté interesado en analizar el discurso de esta obra desde un punto de vista lingüístico.

Un primer problema que cabe mencionar son los propios criterios de edición utilizados. En la recopilación de datos para grandes corpus históricos, no es usual crear las ediciones que van a formar parte del corpus, sino que, por razones prácticas, se recurre a ediciones existentes, independientemente de los criterios utilizados. El problema de esta metodología es que las ediciones empleadas no siempre son las más adecuadas para explotar lingüísticamente los datos, pues se han realizado con fines diferentes, como facilitar la lectura del texto. En el caso que nos ocupa, es posible rastrear ejemplos que reflejan una intervención ortográfica del editor, lo cual limita las posibilidades de análisis lingüístico. Por ejemplo, al corregirse las formas propias del judeoespañol *moyca* y *moycada* como *mosca* y *moscada*, respectivamente, se elimina un rasgo lingüístico característico de personajes judíos, como son *Judío* y *Lozana* (Díaz-Bravo, 2019: 19). Asimismo, la corrección de *delantre* (mamotreto VIII) por *delante* impide rastrear la existencia de este arcaísmo:

[12] Sobre los problemas que presentan los corpus diacrónicos para estudios lingüísticos, véase Díaz-Bravo (2015, 2018).

Facsímil (BNA)	
LD	yo pareçer delantre a otra que fuera en todo el mundo de velleza y bien quista
CORDE, CDH	yo parecer delante a otra que fuera en todo el mundo, de belleza y bienquista
CdE	yo paresçer delante a otra que fuera en todo el mundo de belleza y bienquista

Tabla 4. Comparación de transcripciones: LD /vs/ corpus diacrónicos del español

Un segundo problema atañe a la presentación del documento. Puesto que los corpus ofrecen una única versión del texto, modernizada en mayor o menor medida, todos los resultados obtenidos se ciñen a esta versión, sin posibilidad de contrastar los datos ni con el documento original (*i.e.*, una edición facsimilar) ni con versiones alternativas que faciliten la obtención de variantes ortográficas del original. En el caso de LD, cualquier usuario puede consultar la versión transcrita o la normalizada y comparar resultados para obtener información de tipo grafemático (*vid.* figura 12 y tabla 3). Asimismo, siempre es posible acudir al documento original a través de la imagen facsimilar. LD facilita una comparación entre texto e imagen mediante la visualización paralela de ambas versiones (*vid.* figura 6), y también permite mostrar el número de línea del texto, para agilizar su localización.

[16] obrina, más ha de los años treynta que io no vi a vuestro padre, porque se fue
[17] niño; y después me dixeron que se casó por amores con vuestra madre, y en vos
[18] veo io que vuestra madre hera hermossa. Loçana: ¿Yo, señora? Pues más paresco
[19] a mi agüela que a mi señora madre; y por amor de mi agüela me llamaron a mí
[20] Aldronça; y si esta mi agüela biuía, sabía yo más que no sé, que ella me mostró

Figura 16. Visualización de un fragmento de texto con numeración de líneas

En tercer lugar, debemos destacar el sistema de búsqueda, que multiplica las opciones de análisis con respecto a los tres corpus citados. Por ejemplo, en una primera búsqueda se pueden obtener todos los sustantivos del corpus y, posteriormente, agrupar la frecuencia por lema. De este modo se obtiene fácilmente la lista de sustantivos más frecuentes usados en RLA (*vid.* tabla 4). No hemos conseguido obtener este tipo de información mediante la interfaz de búsqueda del CdE ni del CDH (excluimos el CORDE por no estar lematizado).

Por último, el sistema de búsquedas permite cruzar información de tipo lingüístico con metadatos. Esta posibilidad resulta particularmente interesante, por ejemplo, para analizar el discurso de un personaje determinado o de un grupo de personajes asociados a una categoría social concreta: hombres/mujeres, analfabetos/cultos… (*vid.* figura 15).

6. Conclusiones

En este trabajo hemos presentado *Lozana Digital* (LD), un recurso electrónico que combina una edición digital del *Retrato de la Lozana andaluza* con un corpus anotado para realizar estudios lingüísticos sobre esta obra. A través de la plataforma TEITOK, que da soporte en línea al recurso aquí presentado, LD se presenta como una herramienta digital pensada tanto para un público lego interesado en la lectura de esta obra, como para un

público especializado e interesado en su estudio. Con relación a este último perfil de usuario, LD aúna metodologías y técnicas propias de las humanidades digitales y de la lingüística de corpus para satisfacer tanto al filólogo atraído por aspectos editoriales como al lingüista que busca obtener estadísticas relativas al uso del lenguaje.

Entre los aspectos que ofrece LD a cualquier usuario, cabe citar los siguientes:

- La presentación de la obra en diferentes modos de visualización: edición conservadora, normalizada y facsimilar.
- La visualización paralela de texto e imagen facsimilar, y la opción de mostrar la numeración de líneas y la paginación para facilitar la localización del texto en la imagen.
- La lista completa de personajes con sus correspondientes fichas informativas, incluyendo la posibilidad de acceder desde la ficha de un personaje a los mamotretos concretos en que ese personaje interviene en la obra.
- La posibilidad de descargar el texto completo en formato XML o en formato TXT.
- La obtención de concordancias y frecuencias de uso de una palabra o expresión determina.
- La posibilidad de buscar por forma transcrita, forma normalizada, etiqueta morfosintáctica y lema.
- La obtención de listas de frecuencias agrupadas en función de diferentes criterios: por ejemplo, la obtención de todos los nombres comunes y su agrupación por lema para consultar los sustantivos que aparecen con mayor o menor frecuencia en el texto.
- La posibilidad de realizar búsquedas cruzadas de aspectos lingüísticos y extralingüísticos: por ejemplo, la búsqueda de una expresión o un patrón morfosintáctico concretos para comparar su frecuencia de uso en personajes cultos/analfabetos, o femeninos/masculinos.

En definitiva, *Lozana Digital* constituye una herramienta útil para realizar estudios lingüísticos, tanto cuantitativos como cualitativos, del *Retrato de la Lozana andaluza* y, por tanto, para analizar en profundidad el discurso de esta obra singular de la literatura española quinientista. Esperamos con ello que este recurso sea una aportación provechosa para la investigación en el campo de la lingüística histórica del español.

7. Referencias Bibliográficas

Allaigre, C. (1985). *La Lozana Andaluza*, edición, introducción y notas de C. Allaigre. Madrid: Cátedra.

Allegra, G. (1983). *La Lozana Andaluza*. Edición, introducción y notas de G. Allegra. Madrid: Taurus.

Anipa, K. (2001). *A Critical Examination of Linguistic Variation in Golden-Age Spanish*. Nueva York, Oxford: Peter Lang.

Bubnova, T. (Ed.) (2008). *Retrato de la Lozana andaluza*, edición de Tatiana Bubnova. Doral: Stockcero.

BVMC (Ed.) (2004) = Biblioteca Virtual Miguel de Cervantes (2004). Edición digital de *La lozana andaluza*, basada en la edición de Damiani (1984 [1969]). En línea: http://www.cervantesvirtual.com/obra-visor/la-lozana-andaluza--1/html/00132f70-82b2-11df-acc7-002185ce6064_2.html#I_0_

---------- (Ed.) (2003) = Biblioteca Virtual Miguel de Cervantes (2003). Edición digital de *La Lozana Andaluza*, basada en la edición de Chiclana (1988). En línea: http://www.cervantesvirtual.com/obra-visor/la-lozana-andaluza--0/html/fedbdd78-82b1-11df-acc7-002185ce6064_1.html#I_1_

Calderón-Campos M. (2019). La edición de corpus lingüísticos en la plataforma TEITOK. El caso de *Oralia diacrónica del español (ODE). Chimera: Romance Corpora and Linguistic Studies*, 6, 21-36.

CdE = Davies, M. *Corpus del Español: Género/Histórico*. En línea: https://www.corpusdelespanol.org/hist-gen/
CDH = Real Academia Española. *Corpus del Nuevo Diccionario Histórico del Español*. En línea: http://web.frl.es/CNDHE/
Chiclana, Á. (1988). *La Lozana andaluza*, edición e introducción de Á. Chiclana. Madrid: Espasa Calpe, Colección Austral.
CORDE = Real Academia Española. *Corpus Diacrónico del Español*. En línea: http://corpus.rae.es/cordenet.html
Damiani, B. (Ed.) (1969). *La Lozana Andaluza*, edición, introducción y notas de B. Damiani. Madrid: Castalia.
Delicado, F. (1530?). *Retrato de la Loçana Andaluza. El qual Retrato demuestra lo que en Roma passava y contiene munchas mas cosas que la Celestina*. Edición facsímil de la Biblioteca Nacional de Austria. En línea: http://digital.onb.ac.at/RepViewer/viewer.faces?doc=DTL_6316301. Signatura 66.G.30.(3).
Díaz-Bravo, R. (Ed.) (2019). *Francisco Delicado, Retrato de la Loçana andaluza: Estudio y edición crítica*. Cambridge: Modern Humanities Research Association.
---------- (2018). Las Humanidades Digitales y los corpus diacrónicos en línea del español: problemas y sugerencias. En E. Romero-Frías y L. Bocanegra-Barbecho (Eds.), *Ciencias Sociales y Humanidades Digitales Aplicadas* (pp. 577-602). Granada/Nueva York: Universidad de Granada/Downhill Publishing.
---------- (2015). Herramientas computacionales aplicadas al estudio de la Historia de la Lengua Española. En J. P. Sánchez-Méndez, M. de La Torre y V. Codita (Coords.), *Temas, problemas y métodos para la edición y el estudio de documentos hispánicos antiguos* (pp. 377-394). Valencia: Tirant lo Blanch.
---------- (2010). *Estudio de la oralidad en el Retrato de la Loçana andaluza (Roma, 1524)*. Málaga: Universidad de Málaga.
---------- & Fernández-Alcaide, M. (2018). La oralidad en el siglo XVI. Lo literario y lo privado (I). Marcadores discursivos. *Bulletin of Hispanic Studies*, 95, 357-381.
---------- & Vaamonde, G. (2019). *LD. Lozana Digital*. En línea: http://corpora.ugr.es/lozana/
Gernert, F. & Joset, J. (Eds.) (2013). *La Lozana andaluza*, edición de F. Gernert & J. Joset. Madrid: Real Academia Española - Barcelona: Galaxia Gutenberg, Círculo de Lectores, Biblioteca Clásica de la Real Academia Española, volumen 22.
Gries, S. T. & Berez, A. L. (2017). Annotation in/for Corpus Linguistics. En N. Ide y J. Pustejovsky (Eds.), *Handbook of Linguistic Annotation* (pp. 379-410). Berlín: Springer.
Janssen, M. (2014). TEITOK. *A Tokenized TEI environment*. http://teitok.corpuswiki.org/site/index.php
---------- (2016). TEITOK. Text-Faithful Annotated Corpora. *Proceedings of the Tenth International Conference on Language resources and Evaluation (LREC 2016)*. Portorož, Slovenia, pp. 4037-4043.
Koch, P. & Oesterreicher, W. (2007). *Lengua hablada en la Romania: español, francés, italiano*, traducido del alemán por A. López Serena. Madrid: Gredos.
Leech, G. & Wilson, A. (1996). *Recommendations for the Morphosyntactic Annotation of Corpora. EAGLES Document EAG-TCWG-MAC/R*, marzo de 1996. Disponible en http://www.ilc.cnr.it/EAGLES96/annotate/annotate.html
Medina-Morales, F. (2005). *La Lengua del Siglo de Oro: un estudio de variación lingüística*. Granada: Universidad de Granada.
Oesterreicher, W. (2004). Textos entre inmediatez y distancia comunicativas. El problema de lo hablado escrito en el Siglo de Oro. En R. Cano-Aguilar (Coord.), *Historia de la lengua española* (pp. 729–769). Barcelona: Ariel.
Pérez-Gómez, A. (Ed.) (1950). *Retrato de la Loçana andaluza, en lengua española, muy claríssima. Compuesto en Roma*, Venecia. Edición facsímil. Valencia: Talleres de Tipografía Moderna. Disponible en línea en la BVMC: http://www.cervantesvirtual.com/obra-visor/retrato-de-la-lozana-andaluza--0/html/ffcfbd8a-82b1-11df-acc7-002185ce6064.htm.
Perugini, C (Ed.) (2004). *La Lozana andaluza*, edición, introducción y notas de C. Perugini. Sevilla: Fundación Lara.
TEI Consortium (Eds.) (2019a). *TEI P5: Guidelines for Electronic Text Encoding and Interchange*. Versión 3.6.0. http://www.tei-c.org/Guidelines/P5/.
---------- (Eds.) (2019b). 13.3.2 The Person Element. *TEI P5: Guidelines for Electronic Text Encoding and Interchange*. Versión 3.6.0. https://www.tei-c.org/release/doc/tei-p5-doc/en/html/ND.html#NDPERSE.

Tempesta, D. & Carocci, M. Glossari di Ispanistica: Glossario della Lozana andaluza. http://cisadu2.let.uniroma1.it/glosarios/lozana/

Vaamonde, G. (2018a). Escritura epistolar, edición digital y anotación de corpus. *Cuadernos del Instituto Historia de la Lengua*, 11, 139-164.

---------- (2018b). La multidisciplinariedad en la creación de corpus históricos: El caso de Post Scriptum. *Artnodes*, Humanidades digitales: sociedades, políticas, saberes, 22, 118-127.